RELATIVE VALUE ITERATION FOR STOCHASTIC DIFFERENTIAL GAMES

ARI ARAPOSTATHIS, VIVEK S. BORKAR, AND K. SURESH KUMAR

ABSTRACT. We study zero-sum stochastic differential games with player dynamics governed by a nondegenerate controlled diffusion process. Under the assumption of uniform stability, we establish the existence of a solution to the Isaac's equation for the ergodic game and characterize the optimal stationary strategies. The data is not assumed to be bounded, nor do we assume geometric ergodicity. Thus our results extend previous work in the literature. We also study a relative value iteration scheme that takes the form of a parabolic Isaac's equation. Under the hypothesis of geometric ergodicity we show that the relative value iteration converges to the elliptic Isaac's equation as time goes to infinity. We use these results to establish convergence of the relative value iteration for risk-sensitive control problems under an asymptotic flatness assumption.

1. Introduction

In this paper we consider a relative value iteration for zero-sum stochastic differential games. This relative value iteration is introduced in [1] for stochastic control, and we follow the method introduced in this paper.

In Section 2, we prove the existence of a solution to the Isaac's equation corresponding to the ergodic zero-sum stochastic differential game. We do not assume that the data or the running payoff function is bounded, nor do we assume geometric ergodicity, so our results extend the work in [4]. In Section 3, we introduce a relative value iteration scheme for the zero-sum stochastic differential game and prove its convergence under a hypothesis of geometric ergodicity. In Section 4, we apply the results from Section 3 and study a value iteration scheme for risk-sensitive control under an asymptotic flatness assumption.

2. Problem Description

We consider zero-sum stochastic differential games with state dynamics modeled by a controlled nondegenerate diffusion process $X = \{X(t) : 0 \le t < \infty\}$, and subject to a long-term average payoff criterion.

2.1. **State dynamics.**

Let U_i, $i = 1, 2$, be compact metric spaces and $V_i = \mathcal{P}(U_i)$ denote the space of all probability measures on U_i with Prohorov topology. Let $\bar{b} : \mathbb{R}^d \times U_1 \times U_2 \to \mathbb{R}^d$

1991 *Mathematics Subject Classification.* Primary, 93E15, 93E20; Secondary, 60J25, 60J60, 90C40.

Ari Arapostathis was supported in part by ONR under the Electric Ship Research and Development Consortium.

Vivek Borkar was supported in part by Grant #11IRCCSG014 from IRCC, IIT, Mumbai.

and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be measurable functions. Assumptions on $\bar{b}$ and σ will be specified later. Define $b : \mathbb{R}^d \times V_1 \times V_2 \to \mathbb{R}^d$ as

$$b(x, v_1, v_2) := \int_{U_1}\int_{U_2} \bar{b}(x, u_1, u_2)\, v_1(\mathrm{d}u_1)\, v_2(\mathrm{d}u_2)\,,$$

for $x \in \mathbb{R}^d$, $v_1 \in V_1$ and $v_2 \in V_2$. We model the controlled diffusion process X via the Itô s.d.e.

$$\mathrm{d}X(t) \;=\; b\bigl(X(t), v_1(t), v_2(t)\bigr)\,\mathrm{d}t + \sigma\bigl(X(t)\bigr)\,\mathrm{d}W(t)\,. \tag{2.1}$$

All processes in (2.1) are defined in a common probability space $(\Omega,\, \mathcal{F},\, P)$ which is assumed to be complete. The process $W = \{W(t) : 0 \le t < \infty\}$ is an $\mathbb{R}^d$-valued standard Wiener process which is independent of the initial condition X_0 of (2.1). Player i, with $i = 1, 2$, controls the dynamics X through her strategy $v_i(\cdot)$, a V_i-valued process which is jointly measurable in $(t, \omega) \in [0, \infty) \times \Omega$ and non-anticipative, i.e., for $s < t$, $W(t) - W(s)$ is independent of

$$\mathcal{F}_s \;:=\; \text{the completion of } \sigma(X_0, v_1(r), v_2(r), W(r), r \le s)\,.$$

We denote the set of all such controls (admissible controls) for player i by $\mathcal{U}_i$, $i = 1, 2$.

Assumptions on the Data: We assume the following conditions on the coefficients $\bar{b}$ and σ to ensure existence of a unique solution to (2.1).

- (A1) The functions $\bar{b}$ and σ are locally Lipschitz continuous in $x \in \mathbb{R}^d$, uniformly over $(u_1, u_2) \in U_1 \times U_2$, and have at most a linear growth rate in $x \in \mathbb{R}^d$. Also $\bar{b}$ is continuous.
- (A2) For each $R > 0$ there exists a constant $\kappa(R) > 0$ such that

$$z^{\mathsf{T}} a(x) z \;\ge\; \kappa(R)\|z\|^2 \qquad \text{for all } \|x\| \le R \text{ and } z \in \mathbb{R}^d\,,$$

  where $a := \sigma\sigma^{\mathsf{T}}$, with $^{\mathsf{T}}$ denoting the transpose.

Definition 2.1. For $f \in C^2(\mathbb{R}^d)$ define

$$\bar{L}f(x, u_1, u_2) := \bar{b}(x, u_1, u_2)\cdot\nabla f(x) \,+\, \frac{1}{2}\,\mathrm{tr}\bigl(a(x)\nabla^2 f(x)\bigr)$$

for $x \in \mathbb{R}^d$ and $(u_1, u_2) \in U_1 \times U_2$. Also define the *relaxed extended controlled generator* L by

$$Lf(x, v_1, v_2) := \int_{U_1}\int_{U_2} Lf(x, u_1, u_2)\, v_1(\mathrm{d}u_1)\, v_2(\mathrm{d}u_2)\,, \quad f \in C^2(\mathbb{R}^d)\,,$$

for $x \in \mathbb{R}^d$ and $(v_1, v_2) \in V_1 \times V_2$.

We denote the set of all stationary Markov strategies of player i by $\mathcal{M}_i$, $i = 1, 2$.

2.2. **Zero-sum ergodic game.** Let $\bar{r} : \mathbb{R}^d \times U_1 \times U_2 \to [0, \infty)$ be a continuous function, which is also locally Lipschitz continuous in its first argument. We define the *relaxed running payoff function* $r : \mathbb{R}^d \times V_1 \times V_2 \to [0, \infty)$ by

$$r(x, v_1, v_2) := \int_{U_1}\int_{U_2} \bar{r}(x, u_1, u_2)\, v_1(\mathrm{d}u_1)\, v_2(\mathrm{d}u_2)\,.$$

Player 1 seeks to maximize the average payoff given by

$$\liminf_{T\to\infty} \frac{1}{T}\, E_x\left[\int_0^T r(X(t), v_1(t), v_2(t))\,\mathrm{d}t\right] \tag{2.2}$$

over all admissible controls $v_1 \in \mathcal{U}_1$, while Player 2 seeks to minimize (2.2) over all $v_2 \in \mathcal{U}_2$. Here E_x is the expectation operator corresponding to the probability measure on the canonical space of the process starting at $X(0) = x$.

Assumptions on Ergodicity: We consider the following ergodicity assumptions:

(A3) There exist a nonnegative inf-compact function $\mathcal{V} \in C^2(\mathbb{R}^d)$ and positive constants k_0, k_1 and c such that

$$\begin{aligned} \bar{L}\mathcal{V}(x, u_1, u_2) \;&\le\; k_0 - 2k_1\mathcal{V}(x)\,, \\ \max_{u_1\in U_1, u_2\in U_2} \bar{r}(x, u_1, u_2) \;&\le\; c\,\mathcal{V}(x) \end{aligned}$$

for all $(u_1, u_2) \in U_1 \times U_2$, and $x \in \mathbb{R}^d$.

(A3′) There exist nonnegative inf-compact functions $\mathcal{V} \in C^2(\mathbb{R}^d)$ and $h \in C(\mathbb{R}^d)$, and positive constants k_0 and c such that

$$\begin{aligned} \bar{L}\mathcal{V}(x, u_1, u_2) \;&\le\; k_0 - h(x)\,, \\ \max_{u_1\in U_1, u_2\in U_2} \bar{r}(x, u_1, u_2) \;&\le\; c\,h(x) \end{aligned}$$

for all $(u_1, u_2) \in U_1 \times U_2$, and $x \in \mathbb{R}^d$. Also,

$$\frac{\max_{u_1\in U_1, u_2\in U_2} \bar{r}(x, u_1, u_2)}{h(x)} \xrightarrow[\|x\|\to\infty]{} 0\,.$$

In this section we use assumption (A3′), while in Section 3 we employ (A3) which is stronger and equivalent to geometric ergodicity in the time-homogeneous Markov case. We start with a theorem which characterizes the value of the game under a discounted infinite horizon criterion. For this we need the following notation: For a continuous function $\mathcal{V}\colon \mathbb{R}^d \to (0,\infty)$, $C_{\mathcal{V}}(\mathbb{R}^d)$ denotes the Banach space of functions in $C(\mathbb{R}^d)$ with norm

$$\|f\|_{\mathcal{V}} := \sup_{x\in\mathbb{R}^d} \left|\frac{f(x)}{\mathcal{V}(x)}\right|\,.$$

Theorem 2.1. *Assume* (A1), (A2) *and* (A3′). *For $\alpha > 0$, there exists a solution $\varphi_\alpha \in C_{\mathcal{V}}(\mathbb{R}^d) \cap C^2(\mathbb{R}^d)$ to the p.d.e.*

$$\begin{aligned} \alpha\psi_\alpha(x) \;&=\; \min_{v_2\in V_2}\,\max_{v_1\in V_1}\,\bigl[L\psi_\alpha(x, v_1, v_2) + r(x, v_1, v_2)\bigr] \\ &=\; \max_{v_1\in V_1}\,\min_{v_2\in V_2}\,\bigl[L\psi_\alpha(x, v_1, v_2) + r(x, v_1, v_2)\bigr] \end{aligned} \tag{2.3}$$

and is characterized by

$$\begin{aligned}\psi_\alpha(x) &= \sup_{v_1\in\mathcal{U}_1}\ \inf_{v_2\in\mathcal{U}_2}\ E_x\biggl[\int_0^\infty \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr]\\ &= \inf_{v_2\in\mathcal{U}_2}\ \sup_{v_1\in\mathcal{U}_1}\ E_x\biggl[\int_0^\infty \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr].\end{aligned}$$

Proof. Let B_R denote the open ball of radius R centered at the origin in $\mathbb{R}^d$. The p.d.e.

$$\begin{aligned}\alpha\varphi_\alpha^R(x) &= \min_{v_2\in V_2}\ \max_{v_1\in V_1}\ \bigl[L\varphi_\alpha^R(x,v_1,v_2)+r(x,v_1,v_2)\bigr]\,,\\ \varphi_\alpha^R &= 0 \qquad \text{on } \partial B_R\end{aligned} \tag{2.4}$$

has a unique solution φ_α^R in $C^2(B_R)\cap C(\overline{B_R})$, see [6, Theorem 15.12, p. 382]. Since

$$\min_{v_2\in V_2}\ \max_{v_1\in V_1}\ \bigl[L\varphi_\alpha^R(x,v_1,v_2)+r(x,v_1,v_2)\bigr] = \max_{v_1\in V_1}\ \min_{v_2\in V_2}\ \bigl[L\varphi_\alpha^R(x,v_1,v_2)+r(x,v_1,v_2)\bigr]\,,$$

it follows that $\varphi_\alpha^R\in C^2(B_R)\cap C(\overline{B_R})$ is also a solution to

$$\begin{aligned}\alpha\varphi_\alpha^R(x) &= \max_{v_1\in V_1}\ \min_{v_2\in V_2}\ \bigl[L\varphi_\alpha^R(x,v_1,v_2)+r(x,v_1,v_2)\bigr]\,,\\ \varphi_\alpha^R &= 0 \qquad \text{on } \partial B_R\,.\end{aligned} \tag{2.5}$$

Let $v_{1\alpha}^R\colon B_R\to V_1$ be a measurable selector for the maximizer in (2.5) and $v_{2\alpha}^R\colon B_R\to V_2$ be a measurable selector for the minimizer in (2.4). Then the p.d.e.

$$\begin{aligned}\alpha\varphi_\alpha^R(x) &= \min_{v_2\in V_2}\ \bigl[L\varphi_\alpha^R(x,v_{1\alpha}^R(x),v_2)+r(x,v_{1\alpha}^R(x),v_2)\bigr]\,,\\ \varphi_\alpha^R &= 0 \qquad \text{on } \partial B_R\end{aligned}$$

has a unique solution $\varphi_\alpha^R\in C^{2,r}(B_R)\cap C(\overline{B_R})$, $0<r<1$. By a routine application of Itô's formula, it follows that

$$\varphi_\alpha^R(x) = \inf_{v_2\in\mathcal{U}_2}\ E_x\biggl[\int_0^{\tau_R} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_{1\alpha}^R(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,, \tag{2.6}$$

where

$$\tau_R := \inf\,\{t\ge 0 : \|X(t)\|\ge R\}$$

and X is the solution to (2.1) corresponding to the control pair $(v_{1\alpha}^R, v_2)$, with $v_2\in\mathcal{U}_2$.

Repeating the above argument with the outer minimizer $v_{2\alpha}^R$ of (2.4), we similarly obtain

$$\varphi_\alpha^R(x) = \sup_{v_1\in\mathcal{U}_1}\ E_x\biggl[\int_0^{\tau_R} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_{2\alpha}^R(X(t))\bigr)\,\mathrm{d}t\biggr]\,. \tag{2.7}$$

Combining (2.6) and (2.7), we obtain

$$\begin{aligned}\inf_{v_2\in\mathcal{U}_2}\ \sup_{v_1\in\mathcal{U}_1}\ E_x\biggl[\int_0^{\tau_R} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr] &\le \varphi_\alpha^R(x)\\ &\le \sup_{v_1\in\mathcal{U}_1}\ \inf_{v_2\in\mathcal{U}_2}\ E_x\biggl[\int_0^{\tau_R} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,,\end{aligned}$$

which implies that

$$\begin{aligned}\varphi_\alpha^R(x) &= \sup_{v_1\in\mathcal{U}_1}\ \inf_{v_2\in\mathcal{U}_2}\ E_x\biggl[\int_0^{\tau_R} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr]\\ &= \inf_{v_2\in\mathcal{U}_2}\ \sup_{v_1\in\mathcal{U}_1}\ E_x\biggl[\int_0^{\tau_R} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr].\end{aligned}$$

It is evident that $\varphi_\alpha^R(x) \le \tilde\psi_\alpha(x)$, $x\in\mathbb{R}^d$, where

$$\tilde\psi_\alpha(x) := \sup_{v_1\in\mathcal{U}_1}\ \inf_{v_2\in\mathcal{U}_2}\ E_x\biggl[\int_0^{\infty} \mathrm{e}^{-\alpha t} r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr],\quad x\in\mathbb{R}^d.$$

Also φ_α^R is nondecreasing in R. By Assumption (A3$'$), it follows that

$$\tilde\psi_\alpha(x) \le c\,E_x\biggl[\int_0^{\infty} \mathrm{e}^{-\alpha t} h\bigl(X(t)\bigr)\,\mathrm{d}t\biggr],$$

where X is a solution to (2.1) corresponding to some stationary Markov control pair. Since the function $x\mapsto E_x\bigl[\int_0^\infty \mathrm{e}^{-\alpha t}h\bigl(X(t)\bigr)\,\mathrm{d}t\bigr]$ is continuous, it follows that $\tilde\psi_\alpha\in L^p_{loc}(\mathbb{R}^d)$ for $1<p<\infty$.

By Beneš' measurable selection theorem [3] there exists a pair of controls $(v_{1\alpha}^R, v_{2\alpha}^R)\in\mathcal{M}_1\times\mathcal{M}_2$ which realizes the minimax in (2.4)–(2.5), i.e., for all $x\in B_R$ the following holds:

$$\max_{v_1\in V_1}\ \min_{v_2\in V_2}\ \bigl[L\varphi_\alpha^R(x,v_1,v_2)+r(x,v_1,v_2)\bigr] = L\varphi_\alpha^R\bigl(x,v_{1\alpha}^R(x),v_{2\alpha}^R(x)\bigr)+r\bigl(x,v_{1\alpha}^R(x),v_{2\alpha}^R(x)\bigr).$$

Hence $\varphi_\alpha^R\in C^2(B_R)\cap C(\overline{B_R})$ is a solution to

$$\alpha\varphi_\alpha^R(x) = L\varphi_\alpha^R\bigl(x,v_{1\alpha}^R(x),v_{2\alpha}^R(x)\bigr)+r\bigl(x,v_{1\alpha}^R(x),v_{2\alpha}^R(x)\bigr),\quad x\in B_R.$$

Hence by [2, Lemma A.2.5, p. 305], for each $1<p<\infty$ and $R'>2R$, we have

$$\begin{aligned}\bigl\|\varphi_\alpha^{R'}\bigr\|_{W^{2,p}(B_R)} &\le K_1\Bigl(\bigl\|\varphi_\alpha^{R'}\bigr\|_{L^p(B_{2R})}+\bigl\|L\varphi_\alpha^{R'}-\alpha\varphi_\alpha^{R'}\bigr\|_{L^p(B_{2R})}\Bigr)\\ &\le K_1\Bigl(\bigl\|\tilde\psi_\alpha\bigr\|_{L^p(B_{2R})}+\bigl\|r\bigl(\,\cdot\,,v_{1\alpha}^{R'}(\,\cdot\,),v_{2\alpha}^{R'}(\,\cdot\,)\bigr)\bigr\|_{L^p(B_{2R})}\Bigr)\\ &\le K_1\Bigl(\bigl\|\tilde\psi_\alpha\bigr\|_{L^p(B_{2R})}+K_2(R)|B_{2R}|^{1/p}\Bigr),\end{aligned}$$

where $K_1>0$ is a constant independent of R' and $K_2(R)$ is a constant depending only on the bound of r on B_{2R}. Using standard approximation arguments involving Sobolev imbedding theorems, see [2, p. 111], it follows that there exists $\psi_\alpha\in W^{2,p}_{loc}(\mathbb{R}^d)$ such that $\varphi_\alpha^R\uparrow\psi_\alpha$ as $R\uparrow\infty$ and ψ_α is a solution to

$$\alpha\psi_\alpha(x) = \max_{v_1\in V_1}\ \min_{v_2\in V_2}\ \bigl[L\psi_\alpha(x,v_1,v_2)+r(x,v_1,v_2)\bigr].$$

By standard regularity arguments, see [2, p. 109], one can show that $\psi_\alpha\in C^{2,r}(\mathbb{R}^d)$, $0<r<1$. Also using the minimax condition, it follows that $\psi_\alpha\in C^{2,r}(\mathbb{R}^d)$, $0<r<1$, is a

solution to

$$\begin{aligned}\alpha\psi_\alpha(x) &= \min_{v_2\in V_2}\max_{v_1\in V_1}\bigl[L\psi_\alpha(x,v_1,v_2)+r(x,v_1,v_2)\bigr]\\ &= \max_{v_1\in V_1}\min_{v_2\in V_2}\bigl[L\psi_\alpha(x,v_1,v_2)+r(x,v_1,v_2)\bigr]\,.\end{aligned}$$

Let $v_1^\alpha\in\mathcal{M}_1$ and $v_2^\alpha\in\mathcal{M}_2$ be an outer maximizing and an outer minimizing selector for (2.3), respectively, corresponding to ψ_α given above. Then ψ_α satisfies the p.d.e.

$$\alpha\psi_\alpha(x) = \max_{v_1\in V_1}\bigl[L\psi_\alpha\bigl(x,v_1,v_2^\alpha(x)\bigr)+r\bigl(x,v_1,v_2^\alpha(x)\bigr)\bigr]\,.$$

For $v_1\in\mathcal{U}_1$, let X be the solution to (2.1) corresponding to (v_1,v_2^α) and the initial condition $x\in\mathbb{R}^d$. Applying the Itô–Dynkin formula, we obtain

$$E_x\bigl[\mathrm{e}^{-\alpha\tau_R}\psi_\alpha(X(\tau_R))\bigr]-\psi_\alpha(x) \le -E_x\biggl[\int_0^{\tau_R}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1(t),v_2^\alpha(X(t))\bigr)\,\mathrm{d}t\biggr]\,.$$

Since $\psi_\alpha\ge 0$, we have

$$\psi_\alpha(x) \ge E_x\biggl[\int_0^{\tau_R}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1(t),v_2^\alpha(X(t))\bigr)\,\mathrm{d}t\biggr]\,.$$

Using Fatou's lemma we obtain

$$\psi_\alpha(x) \ge E_x\biggl[\int_0^{\infty}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1(t),v_2^\alpha(X(t))\bigr)\,\mathrm{d}t\biggr]\,. \tag{2.8}$$

Therefore

$$\psi_\alpha(x) \ge \sup_{v_1\in\mathcal{U}_1} E_x\biggl[\int_0^{\infty}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1(t),v_2^\alpha(X(t))\bigr)\,\mathrm{d}t\biggr]\,. \tag{2.9}$$

Similarly, for $v_2\in\mathcal{U}_2$, let X be the solution to (2.1) corresponding to (v_1^α,v_2) and the initial condition $x\in\mathbb{R}^d$. By applying the Itô–Dynkin formula, we obtain

$$E_x\bigl[\mathrm{e}^{-\alpha\tau_R}\psi_\alpha(X(\tau_R))\bigr]-\psi_\alpha(x) \ge -E_x\biggl[\int_0^{\tau_R}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1^\alpha(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,.$$

Hence

$$\psi_\alpha(x) \le E_x\biggl[\int_0^{\infty}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1^\alpha(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr] + E_x\bigl[\mathrm{e}^{-\alpha\tau_R}\psi_\alpha(X(\tau_R))\bigr]\,.$$

By [2, Remark A.3.8, p. 310], it follows that

$$\lim_{R\uparrow\infty} E_x\bigl[\mathrm{e}^{-\alpha\tau_R}\psi_\alpha(X(\tau_R))\bigr] = 0\,.$$

Hence, we have

$$\psi_\alpha(x) \le E_x\biggl[\int_0^{\infty}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1^\alpha(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,. \tag{2.10}$$

Therefore

$$\psi_\alpha(x) \le \inf_{v_2\in\mathcal{U}_2} E_x\biggl[\int_0^{\infty}\mathrm{e}^{-\alpha t}r\bigl(X(t),v_1^\alpha(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,. \tag{2.11}$$

By (2.9) and (2.11), we obtain

$$\psi_\alpha(x) \;=\; E_x\biggl[\int_0^\infty \mathrm{e}^{-\alpha t} r\bigl(X(t), v_1^\alpha(X(t)), v_2^\alpha(X(t))\bigr)\,\mathrm{d}t\biggr]. \tag{2.12}$$

Also by (2.8) and (2.10) we have

$$\begin{aligned}\inf_{v_2\in\mathcal{U}_2}\,\sup_{v_1\in\mathcal{U}_1}\; E_x\biggl[\int_0^\infty \mathrm{e}^{-\alpha t} r\bigl(X(t), v_1(t), v_2(t)\bigr)\,\mathrm{d}t\biggr] \;&\le\; \psi_\alpha(x)\\ &\le\; \sup_{v_1\in\mathcal{U}_1}\,\inf_{v_2\in\mathcal{U}_2}\; E_x\biggl[\int_0^\infty \mathrm{e}^{-\alpha t} r\bigl(X(t), v_1(t), v_2(t)\bigr)\,\mathrm{d}t\biggr].\end{aligned}$$

This implies the desired characterization. □

Remark 2.1. Using Theorem 2.1, one can easily show that any pair of measurable outer maximizing and outer minimizing selectors of (2.3) is a saddle point equilibrium for the stochastic differential game with state dynamics given by (2.1) and with a discounted criterion under the running payoff function r.

Theorem 2.2. *Assume* (A1), (A2) *and* (A3′). *Then there exists a solution* $(\beta, \varphi^*) \in \mathbb{R}\times C_{\mathcal{V}}(\mathbb{R}^d)\cap C^2(\mathbb{R}^d)$ *to the Isaac's equation*

$$\begin{aligned}\beta \;&=\; \min_{v_2\in V_2}\,\max_{v_1\in V_1}\,\bigl[L\varphi^*(x,v_1,v_2) + r(x,v_1,v_2)\bigr]\\ &=\; \max_{v_1\in V_1}\,\min_{v_2\in V_2}\,\bigl[L\varphi^*(x,v_1,v_2) + r(x,v_1,v_2)\bigr],\\ \varphi^*(0) \;&=\; 0\end{aligned} \tag{2.13}$$

such that β *is the value of the game.*

Proof. For $(v_1, v_2)\in\mathcal{M}_1\times\mathcal{M}_2$, define

$$J_\alpha(x, v_1, v_2) := E_x\biggl[\int_0^\infty \mathrm{e}^{-\alpha t} r\bigl(X(t), v_1(X(t)), v_2(X(t))\bigr)\,\mathrm{d}t\biggr], \quad x\in\mathbb{R}^d\,,$$

where X is a solution to (2.1) corresponding to $(v_1, v_2)\in\mathcal{M}_1\times\mathcal{M}_2$. Hence from (2.12), we have

$$\psi_\alpha(x) \;=\; J_\alpha(x, v_1^\alpha, v_2^\alpha)\,,$$

where $(v_1^\alpha, v_2^\alpha)\in\mathcal{M}_1\times\mathcal{M}_2$ is a pair of measurable outer maximizing and outer minimizing selectors of (2.3). Using (A3′), it is easy to see that (v_1^α, v_2^α) is a pair of stable stationary Markov controls. Hence by the arguments in the proof of [2, Theorem 3.7.4, pp. 128–131], we have the following estimates:

$$\begin{aligned}\|\psi_\alpha - \psi_\alpha(0)\|_{W^{2,p}(B_R)} \;\le\; \frac{K_3}{\eta[v_1^\alpha, v_2^\alpha](B_R)}\Biggl(&\frac{\beta[v_1^\alpha, v_2^\alpha]}{\eta[v_1^\alpha, v_2^\alpha](B_R)}\\ &+ \max_{(x,v_1,v_2)\in B_{4R}\times V_1\times V_2} r(x,v_1,v_2)\Biggr),\end{aligned} \tag{2.14}$$

$$\sup_{x\in B_R)} \alpha\psi_\alpha(x) \;\le\; K_3\Biggl(\frac{\beta[v_1^\alpha, v_2^\alpha]}{\eta[v_1^\alpha, v_2^\alpha](B_R)} + \max_{(x,v_1,v_2)\in B_{4R}\times V_1\times V_2} r(x,v_1,v_2)\Biggr), \tag{2.15}$$

where $\eta[v_1^\alpha, v_2^\alpha]$ is the unique invariant probability measure of the process (2.1) corresponding to (v_1^α, v_2^α) and

$$\beta[v_1^\alpha, v_2^\alpha] := \int_{\mathbb{R}^d} r\bigl(x, v_1^\alpha(x), v_2^\alpha(x)\bigr)\, \eta[v_1^\alpha, v_2^\alpha](\mathrm{d}x)\,.$$

It follows from [2, Corollary 3.3.2, p. 97] that

$$\sup_{\alpha>0}\, \beta[v_1^\alpha, v_2^\alpha] \;<\; \infty\,. \tag{2.16}$$

Also from [2, (2.6.9a); p. 69 and (3.3.9); p. 97] it follows that

$$\inf_{\alpha>0}\, \eta[v_1^\alpha, v_2^\alpha](B_R) \;>\; 0\,. \tag{2.17}$$

Combining (2.14)–(2.17), we have

$$\begin{aligned} \|\psi_\alpha - \psi_\alpha(0)\|_{W^{2,p}(B_R)} \;&\le\; K_4\,,\\ \sup_{x\in B_R}\, \alpha\psi_\alpha(x) \;&\le\; K_4\,, \end{aligned} \tag{2.18}$$

where $K_4 > 0$ is a constant independent of $\alpha > 0$.

Define

$$\bar\psi_\alpha(x) := \psi_\alpha(x) - \psi_\alpha(0)\,, \quad x\in\mathbb{R}^d\,.$$

In view of (2.18), one can use the arguments in [2, Lemma 3.5.4, pp. 108–109] to show that along some sequence $\alpha_n \downarrow 0$, $\alpha_n\psi_\alpha(0)$ converges to a constant ϱ and $\bar\psi_{\alpha_n}$ converges uniformly on compact sets to a function $\hat\psi \in C^2(\mathbb{R}^d)$, where the pair $(\varrho,\, \hat\psi)$ is a solution to the p.d.e.

$$\begin{aligned} \varrho \;&=\; \min_{v_2\in V_2}\, \max_{v_1\in V_1}\, \bigl[L\hat\psi(x, v_1, v_2) + r(x, v_1, v_2)\bigr]\,,\\ \hat\psi(0) \;&=\; 0\,. \end{aligned}$$

Moreover, using the Isaac's condition, it follows that $(\varrho, \hat\psi) \in \mathbb{R}\times C^2(\mathbb{R}^d)$ satisfies (2.13).

We claim that $\hat\psi \in o(\mathcal{V})$, i.e., $\frac{\hat\psi(x)}{\mathcal{V}(x)} \to 0$ as $\|x\|\to\infty$. To prove the claim let $(\hat v_1, \hat v_2) \in \mathcal{M}_1\times\mathcal{M}_2$ be a pair of measurable outer maximizing and outer minimizing selectors of (2.13) corresponding to $\hat\psi$. Let X be the solution to (2.1) under the control $(\hat v_1, \hat v_2)$. Then by an application of the Itô–Dynkin formula and the help of Fatou's lemma, we can show that for all $x\in\mathbb{R}^d$

$$\hat\psi(x) \;\ge\; E_x\biggl[\int_0^{\breve\tau_r} \Bigl(r\bigl(X(t), \hat v_1(X(t)), \hat v_2(X(t))\bigr) - \varrho\Bigr)\,\mathrm{d}t\biggr] + \min_{\|y\|=r} \hat\psi(y)\,, \tag{2.19}$$

where

$$\breve\tau_r \;=\; \inf\,\{t\ge 0 : \|X(t)\| \le r\}\,.$$

Let $v_1^\alpha \in \mathcal{M}_1$ be a measurable outer maximizing selector in (2.3). Then the function $\psi_\alpha \in C^{2,r}(\mathbb{R}^d)$ given in Theorem 2.1 satisfies the p.d.e.

$$\alpha\psi_\alpha \;=\; \min_{v_2\in V_2}\, \bigl[L\psi_\alpha(x, v_1^\alpha(x), v_2) + r(x, v_1^\alpha(x), v_2)\bigr]\,. \tag{2.20}$$

Let X be the solution to (2.1) under the control (v_1^α, v_2), with $v_2 \in \mathcal{U}_2$, and initial condition $x \in \mathbb{R}^d$. Then by applying the Itô–Dynkin formula to $\mathrm{e}^{-\alpha t}\psi_\alpha(X(t))$ and using (2.20), we obtain

$$E_x\bigl[\mathrm{e}^{-\alpha(\breve{\tau}_r\wedge\tau_R)}\psi_\alpha(X(\breve{\tau}_r\wedge\tau_R))\bigr] - \psi_\alpha(x) \geq\ -E_x\left[\int_0^{\breve{\tau}_r\wedge\tau_R} r\bigl(X(t), v_1^\alpha(X(t)), v_2(t)\bigr)\,\mathrm{d}t\right],$$

which we write as

$$\psi_\alpha(x)\ \leq\ E_x\left[\int_0^{\breve{\tau}_r} r\bigl(X(t), v_1^\alpha(X(t)), v_2(t)\bigr)\,\mathrm{d}t\right] + E_x\bigl[\mathrm{e}^{-\alpha(\breve{\tau}_r\wedge\tau_R)}\psi_\alpha(X(\breve{\tau}_r\wedge\tau_R))\bigr]\,. \tag{2.21}$$

Using [2, Remark A.3.8, p. 310], it follows that

$$E_x\bigl[\mathrm{e}^{-\alpha\tau_R}\psi_\alpha(X(\tau_R))I\{\breve{\tau}_r\geq\tau_R\}\bigr]\ \leq\ E_x\bigl[\mathrm{e}^{-\alpha\tau_R}\psi_\alpha(X(\tau_R))\bigr] \xrightarrow[R\to\infty]{} 0\,. \tag{2.22}$$

Hence from (2.21) and (2.22), we obtain

$$\psi_\alpha(x)\ \leq\ E_x\left[\int_0^{\breve{\tau}_r} r\bigl(X(t), v_1^\alpha(X(t)), v_2(t)\bigr)\,\mathrm{d}t\right] + E_x\bigl[\mathrm{e}^{-\alpha\breve{\tau}_r}\psi_\alpha(X(\breve{\tau}_r))\bigr]\,.$$

Therefore,

$$\begin{aligned}
\bar{\psi}_\alpha(x)\ &\leq\ E_x\left[\int_0^{\breve{\tau}_r} r\bigl(X(t), v_1^\alpha(X(t)), v_2(t)\bigr)\,\mathrm{d}t\right] + E_x\bigl[\mathrm{e}^{-\alpha\breve{\tau}_r}\psi_\alpha(X(\breve{\tau}_r)) - \psi_\alpha(0)\bigr]\\
&=\ E_x\left[\int_0^{\breve{\tau}_r}\Bigl(r\bigl(X(t), v_1^\alpha(X(t)), v_2(t)\bigr) - \varrho\Bigr)\,\mathrm{d}t\right] + E_x\bigl[\psi_\alpha(X(\breve{\tau}_r)) - \psi_\alpha(0)\bigr]\\
&\qquad\qquad + E_x\bigl[\alpha^{-1}(1-\mathrm{e}^{-\alpha\breve{\tau}_r})\bigl(\varrho - \alpha\psi_\alpha(X(\breve{\tau}_r))\bigr)\bigr]\\
&\leq\ E_x\left[\int_0^{\breve{\tau}_r}\Bigl(r\bigl(X(t), v_1^\alpha(X(t)), v_2(t)\bigr) - \varrho\Bigr)\,\mathrm{d}t\right]\\
&\qquad\qquad + M(r) + E_x\bigl[\breve{\tau}_r\bigr]\sup_{\|y\|=r}\bigl|\varrho - \alpha\psi_\alpha(y)\bigr|\\
&\leq\ \sup_{v_1\in\mathcal{M}_1} E_x\left[\int_0^{\breve{\tau}_r}\Bigl(r\bigl(X(t), v_1(X(t)), v_2(t)\bigr) - \varrho\Bigr)\,\mathrm{d}t\right]\\
&\qquad\qquad + M(r) + \sup_{\|y\|=r}\bigl|\varrho - \alpha\psi_\alpha(y)\bigr| \sup_{v_1\in\mathcal{M}_1} E_x[\breve{\tau}_r]
\end{aligned}$$

for some nonnegative constant $M(r)$ such that $M(r)\to 0$ as $r\downarrow 0$. Next from the definition of $\hat{\psi}$, by letting $\alpha\downarrow 0$ along the sequence given in the proof of Theorem 2.2, we obtain

$$\hat{\psi}(x)\ \leq\ \sup_{v_1\in\mathcal{M}_1} E_x\left[\int_0^{\breve{\tau}_r}\Bigl(r\bigl(X(t), v_1(X(t)), v_2(t)\bigr) - \varrho\Bigr)\,\mathrm{d}t\right] + M(r)\,. \tag{2.23}$$

By combining (2.19) and (2.23), the result follows by [2, Lemma 3.7.2, p. 125]. This completes the proof of the claim.

Let $(\hat{v}_1, \hat{v}_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ be a pair of measurable outer maximizing and minimizing selectors in (2.13) corresponding to $\hat{\psi}$. Then $(\varrho, \hat{\psi})$ satisfies the p.d.e.

$$\varrho \;=\; \max_{v_1 \in V_1} \left[L\hat{\psi}\bigl(x, v_1, \hat{v}_2(x)\bigr) + r\bigl(x, v_1, \hat{v}_2(x)\bigr) \right].$$

Let $v_1 \in \mathcal{U}_1$ and X be the process in (2.1) under the control $(v_1, \hat{v}_2)$ and initial condition $x \in \mathbb{R}^d$. By applying the Itô–Dynkin formula, we obtain

$$E_x\bigl[\hat{\psi}(X(t \wedge \tau_R))\bigr] - \hat{\psi}(x) \;\le\; -E_x\left[\int_0^{t\wedge\tau_R} \Bigl(r\bigl(X(t), v_1(t), \hat{v}_2(X(t))\bigr) - \varrho\Bigr)\,\mathrm{d}t\right].$$

Hence

$$\varrho t \;\ge\; E_x\left[\int_0^{t\wedge\tau_R} r\bigl(X(t), v_1(t), \hat{v}_2(X(t))\bigr)\,\mathrm{d}t\right] + E_x\bigl[\hat{\psi}(X(t \wedge \tau_R))\bigr] - \hat{\psi}(x)$$

for all $t \ge 0$. Using Fatou's lemma and [2, Lemma 3.7.2, p. 125], we obtain

$$\varrho t \;\ge\; E_x\left[\int_0^{t} r\bigl(X(t), v_1(t), \hat{v}_2(X(t))\bigr)\,\mathrm{d}t\right] + E_x\bigl[\hat{\psi}(X(t))\bigr] - \hat{\psi}(x)\,, \quad t \ge 0\,.$$

Dividing by t and taking limits again using [2, Lemma 3.7.2, p. 125], we obtain

$$\varrho \;\ge\; \liminf_{t\to\infty}\; \frac{1}{t}\, E_x\left[\int_0^{t} r\bigl(X(t), v_1(t), \hat{v}_2(X(t))\bigr)\,\mathrm{d}t\right].$$

Since $v_1 \in \mathcal{U}_1$ was arbitrary, we have

$$\begin{aligned} \varrho \;&\ge\; \sup_{v_1\in\mathcal{U}_1}\; \liminf_{t\to\infty}\; \frac{1}{t}\, E_x\left[\int_0^{t} r\bigl(X(t), v_1(t), \hat{v}_2(X(t))\bigr)\,\mathrm{d}t\right] \\ &\ge\; \inf_{v_2\in\mathcal{U}_2}\; \sup_{v_1\in\mathcal{U}_1}\; \liminf_{t\to\infty}\; \frac{1}{t}\, E_x\left[\int_0^{t} r\bigl(X(t), v_1(t), v_2(t)\bigr)\,\mathrm{d}t\right]. \end{aligned} \tag{2.24}$$

The pair $(\varrho, \hat{\psi})$ also satisfies the p.d.e.

$$\varrho \;=\; \min_{v_2 \in V_2} \left[L\hat{\psi}(x, \hat{v}_1(x), v_2) + r(x, \hat{v}_1(x), v_2) \right].$$

Let $v_2 \in \mathcal{U}_2$ and X be the process in (2.1) corresponding to $(\hat{v}_1, v_2)$ and initial condition $x \in \mathbb{R}^d$. By applying the Itô–Dynkin formula, we obtain

$$E_x\bigl[\hat{\psi}(X(t \wedge \tau_R))\bigr] - \hat{\psi}(x) \;\ge\; -E_x\left[\int_0^{t\wedge\tau_R} \Bigl(r\bigl(X(t), \hat{v}_1(X(t)), v_2(t)\bigr) - \varrho\Bigr)\,\mathrm{d}t\right].$$

Hence

$$\varrho E_x[t \wedge \tau_R] \;\le\; E_x\left[\int_0^{t} r\bigl(X(t), \hat{v}_1(X(t)), v_2(t)\bigr)\,\mathrm{d}t\right] + E_x\bigl[\hat{\psi}(X(t \wedge \tau_R))\bigr] - \hat{\psi}(x)\,.$$

Next, by letting $R \to \infty$ and using the dominated convergence theorem for the l.h.s. and [2, Lemma 3.7.2, p. 125] for the r.h.s., we obtain

$$\varrho t \;\le\; E_x\left[\int_0^{t} r\bigl(X(t), \hat{v}_1(X(t)), v_2(t)\bigr)\,\mathrm{d}t\right] + E_x\bigl[\hat{\psi}(X(t))\bigr] - \hat{\psi}(x)\,.$$

Also by [2, Lemma 3.7.2, p. 125], we obtain

$$\varrho \;\le\; \liminf_{t\to\infty}\;\frac{1}{t}\;E_x\biggl[\int_0^t r\bigl(X(t),\hat{v}_1(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,.$$

Since $v_2 \in \mathcal{U}_2$ was arbitrary, we have

$$\begin{aligned}\varrho \;&\le\; \inf_{v_2\in\mathcal{U}_2}\;\liminf_{t\to\infty}\;\frac{1}{t}\;E_x\biggl[\int_0^t r\bigl(X(t),\hat{v}_1(X(t)),v_2(t)\bigr)\,\mathrm{d}t\biggr]\\ &\le\; \sup_{v_1\in\mathcal{U}_1}\;\inf_{v_2\in\mathcal{U}_2}\;\liminf_{t\to\infty}\;\frac{1}{t}\;E_x\biggl[\int_0^t r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,. \qquad (2.25)\end{aligned}$$

Combining (2.24) and (2.25), we obtain

$$\begin{aligned}\varrho \;&=\; \inf_{v_2\in\mathcal{U}_2}\;\sup_{v_1\in\mathcal{U}_1}\;\liminf_{t\to\infty}\;\frac{1}{t}\;E_x\biggl[\int_0^t r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr]\\ &=\; \sup_{v_1\in\mathcal{U}_1}\;\inf_{v_2\in\mathcal{U}_2}\;\liminf_{t\to\infty}\;\frac{1}{t}\;E_x\biggl[\int_0^t r\bigl(X(t),v_1(t),v_2(t)\bigr)\,\mathrm{d}t\biggr]\,,\end{aligned}$$

i.e. $\varrho = \beta$, the value of the game. This completes the proof. □

Remark 2.2. Using Theorem 2.2, one can easily prove that any pair of measurable outer maximizing and outer minimizing selectors of (2.3) is a saddle point equilibrium for the stochastic differential game with state dynamics given by (2.1).

3. Relative Value Iteration

We consider the following relative value iteration equation.

$$\begin{aligned}\frac{\partial\varphi}{\partial t}(t,x) \;&=\; \min_{v_2\in V_2}\;\max_{v_1\in V_1}\;\bigl[L\varphi(t,x,v_1,v_2)+r(x,v_1,v_2)\bigr]\;-\;\varphi(t,0)\;,\\ \varphi(0,x) \;&=\; \varphi_0(x)\;,\end{aligned} \qquad (3.1)$$

where $\varphi_0 \in C_{\mathcal{V}}(\mathbb{R}^d)\cap C^2(\mathbb{R}^d)$. This can be viewed as a continuous time continuous state space variant of the relative value iteration algorithm for Markov decision processes [7].

Convergence of this relative value iteration scheme is obtained through the study of the value iteration equation which takes the form

$$\begin{aligned}\frac{\partial\bar{\varphi}}{\partial t}(t,x) \;&=\; \min_{v_2\in V_2}\;\max_{v_1\in V_1}\;\bigl[L\bar{\varphi}(t,x,v_1,v_2)+r(x,v_1,v_2)\bigr]\;-\;\beta\;,\\ \bar{\varphi}(0,x) \;&=\; \varphi_0(x)\;,\end{aligned} \qquad (3.2)$$

where β is the value of the average payoff game in Theorem 2.2.

Under Assumption (A3), it is straightforward to show that for each $T > 0$ there exists a unique solution $\bar{\varphi}$ in $C_{\mathcal{V}}([0,T]\times\mathbb{R}^d)\cap C^{1,2}([0,T]\times\mathbb{R}^d)$ to the p.d.e. (3.2).

First, we prove the following important estimate which is crucial for the proof of convergence.

Lemma 3.1. *Assume* (A1)–(A3). *Then for each* $T > 0$, *the p.d.e. in* (3.1) *has a unique solution* $\varphi \in C_{\mathcal{V}}([0,T]\times\mathbb{R}^d)\cap C^{1,2}([0,T]\times\mathbb{R}^d)$.

Proof. The proof follows by mimicking the arguments in [1, Lemma 4.1], using the following estimate

$$E_x\bigl[\mathcal{V}(X(t))\bigr] \;\le\; \frac{k_0}{2k_1} + \mathcal{V}(x)\mathrm{e}^{-2k_1 t}\,, \tag{3.3}$$

where X is the solution to (2.1) corresponding to any admissible controls v_1 and v_2 and initial condition $x \in \mathbb{R}^d$. The estimate for φ follows from the arguments in [2, Lemma 2.5.5, pp. 63–64], noting that for all $v_i \in \mathcal{U}_i$, $i = 1, 2$, we have

$$\begin{aligned}\int_0^t E_x\bigl[r^n\bigl(X(s), v_1(s), v_2(s)\bigr)\bigr]\,\mathrm{d}s \;&\le\; c\int_0^t E_x\bigl[\mathcal{V}(X(s))\bigr]\,\mathrm{d}s\\ &\le\; \frac{k_0}{k_1}\bigl(k_0 t + \mathcal{V}(x)\bigr)\,,\end{aligned}$$

where $r^n(x, v_1, v_2) := n \wedge r(x, v_1, v_2)$ is the truncation of r at $n \ge 0$. □

Next, we turn our attention to the p.d.e. in (3.2).

Lemma 3.2. *Assume* (A1)–(A3). *For each $\varphi_0 \in C_{\mathcal{V}}(\mathbb{R}^d) \cap C^2(\mathbb{R}^d)$, the solution $\bar\varphi$ of the p.d.e.* (3.2) *satisfies the following estimate*

$$|\varphi^*(x) \,-\, \bar\varphi(t, x)| \;\le\; \|\varphi^* - \varphi_0\|_{\mathcal{V}}\,\left(\tfrac{k_0}{2k_1} + \mathcal{V}(x)\right) \quad \textit{for all } x \in \mathbb{R}^d\,,\; t \ge 0\,.$$

Proof. Let $v_1^* \in \mathcal{M}_1$ and $v_2^* \in \mathcal{M}_2$ be an outer maximizing and outer minimizing selector of (2.13), respectively. Also let $\bar v_1(\cdot,\cdot)$ and $\bar v_2(\cdot,\cdot)$, respectively, be an outer maximizing and an outer minimizing selector of (3.2). By applying Itô's formula to $\varphi^* - \bar\varphi$, we obtain

$$E_x\bigl[\varphi^*(X^*(t)) - \varphi_0(X^*(t))\bigr] \;\le\; \varphi^*(x) - \bar\varphi(t, x) \;\le\; E_x\bigl[\varphi^*(X(t)) - \varphi_0(X(t))\bigr]\,,$$

where X^*, resp. X is the solution to (2.1) corresponding to (v_1^*, v_2^*) and $(\bar v_1, \bar v_2)$ respectively for the initial condition $x \in \mathbb{R}^d$. An application of (3.3) completes the proof. □

Arguing as in the proof of [1, Lemma 4.4], we can show the following:

Lemma 3.3. *Assume* (A1)–(A3). *If $\bar\varphi(0, x) = \varphi(0, x) = \varphi_0(x)$ for some $\varphi_0 \in C_{\mathcal{V}}(\mathbb{R}^d) \cap C^2(\mathbb{R}^d)$, then*

$$\begin{aligned}\varphi(t, x) - \varphi(t, 0) \;&=\; \bar\varphi(t, x) - \bar\varphi(t, 0)\,,\\ \varphi(t, x) \;&=\; \bar\varphi(t, x) - \mathrm{e}^{-t}\int_0^t \mathrm{e}^{s}\bar\varphi(s, 0)\,\mathrm{d}s + \beta(1 - \mathrm{e}^{-t})\end{aligned}$$

for all $x \in \mathbb{R}^d$ and $t \ge 0$.

Convergence of the relative value iteration is asserted in the following theorem.

Theorem 3.1. *Assume* (A1)–(A3). *For each $\varphi_0 \in C_{\mathcal{V}}(\mathbb{R}^d) \cap C^2(\mathbb{R}^d)$, $\varphi(t, x)$ converges to $\varphi^*(x) + \beta$ as $t \to \infty$.*

Proof. By closely mimicking the arguments of [1, Theorem 4.5], we can show that $\bar\varphi(t, x) \to \varphi^*(x) + c_0$ as $t \to \infty$ for some constant $c_0 \in \mathbb{R}$ which depends on φ_0. By Lemma 3.3 we have

$$\varphi(t, x) = \bar\varphi(t, x) + \int_0^t \mathrm{e}^{s-t}(\beta - \bar\varphi(s, 0))\,\mathrm{d}s\,.$$

Hence $\varphi(t, x) \to \varphi^*(x) + \beta$ as $t \to \infty$. □

4. Risk-Sensitive Control

In this section, we apply the results from Section 3 to study the convergence of a relative value iteration scheme for the risk-sensitive control problem which is described as follows. Let U be a compact metric space and $V = \mathcal{P}(U)$ denote the space of all probability measures on U with Prohorov topology. We consider the risk-sensitive control problem with state equation given by the controlled s.d.e. (in relaxed form)

$$\mathrm{d}X(t) \;=\; b(X(t), v(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t)\,, \tag{4.1}$$

and payoff criterion

$$J(x,v) := \liminf_{T\to\infty}\;\frac{1}{T}\;\ln E\biggl[\exp\biggl(\int_0^T r(X(t),v(t))\,\mathrm{d}t\biggr)\;\Big|\;X(0)=x\biggr]\,.$$

All processes in (4.1) are defined in a common probability space $(\Omega,\,\mathcal{F},\,P)$ which is assumed to be complete. The process W is an $\mathbb{R}^d$-valued standard Wiener process which is independent of the initial condition X_0 of (2.1). The control v is a V-valued process which is jointly measurable in $(t,\omega)\in[0,\infty)\times\Omega$ and non-anticipative, i.e., for $s<t$, $W(t)-W(s)$ is independent of $\mathcal{F}_s :=$ the completion of $\sigma(X_0, v(r), W(r), r\le s)$. We denote the set of all such controls (admissible controls) by $\mathcal{U}$.

Assumptions on the Data: We assume the following properties for the coefficients b and σ:

(B1) The functions b and σ are continuous and bounded, and also Lipschitz continuous in $x\in\mathbb{R}^d$ uniformly over $v\in V$. Also $(\sigma\sigma^{\mathsf{T}})^{-1}$ is Lipschitz continuous.

(B2) For each $R>0$ there exists a constant $\kappa(R)>0$ such that

$$z^{\mathsf{T}}a(x)z \;\ge\; \kappa(R)\|z\|^2 \qquad \text{for all } \|x\|\le R \text{ and } z\in\mathbb{R}^d\,,$$

where $a := \sigma\sigma^{\mathsf{T}}$.

Asymptotic Flatness Hypothesis: We assume the following property:

(B3) (i) There exists a $c>0$ and a positive definite matrix Q such that for all $x,\,y\in\mathbb{R}^d$ with $x\ne y$, we have

$$2\bigl(b(x,v)-b(y,v)\bigr)^{\mathsf{T}}Q(x-y) + \mathrm{tr}\Bigl(\bigl(\sigma(x)-\sigma(y)\bigr)\bigl(\sigma(x)-\sigma(y)\bigr)^{\mathsf{T}}Q\Bigr) - \frac{\bigl\|\bigl(\sigma(x)-\sigma(y)\bigr)^{\mathsf{T}}Q(x-y)\bigr\|^2}{(x-y)^{\mathsf{T}}Q(x-y)} \;\le\; -c\,\|x-y\|^2\,.$$

(ii) Let $\mathrm{Lip}(f)$ denote the Lipschitz constant of a Lipschitz continuous function f. Then

$$2\,\|\sigma\sigma^{\mathsf{T}}\|_\infty^2\,\mathrm{Lip}(r)\,\mathrm{Lip}\bigl((\sigma\sigma^{\mathsf{T}})^{-1}\bigr)\;\le\; c^2.$$

We quote the following result from [5, Theorems 2.2–2.3]:

Theorem 4.1. *Assume* (B1)–(B3). *The p.d.e.*

$$\begin{aligned}\beta &= \min_{v\in V}\,\max_{w\in\mathbb{R}^d}\,\Bigl[\tilde{L}\varphi^*(x,w,v)+r(x,v)-\tfrac{1}{2}\,w^{\mathsf{T}}\bigl(a^{-1}(x)\bigr)w\Bigr]\\ &= \max_{w\in\mathbb{R}^d}\,\min_{v\in V}\,\Bigl[\tilde{L}\varphi^*(x,w,v)+r(x,v)-\tfrac{1}{2}\,w^{\mathsf{T}}\bigl(a^{-1}(x)\bigr)w\Bigr]\,, \qquad (4.2)\\ \varphi^*(0) &= 0\,,\end{aligned}$$

where

$$\tilde{L}f(x,w,v) := (b(x,v)+w)\cdot\nabla f(x)+\frac{1}{2}\operatorname{tr}\bigl(a(x)\nabla^2 f(x)\bigr)\,,\quad f\in C^2(\mathbb{R}^d)\,,$$

has a unique solution $(\beta,\varphi^*)\in\mathbb{R}\times C^2(\mathbb{R}^d)\cap o(\|x\|)$. *Moreover,* β *is the value of the risk-sensitive control problem and any measurable outer minimizing selector in* (4.2) *is risk-sensitive optimal. Also in* (4.2), *the supremum can be restricted to a closed ball* $\tilde{V}=\overline{B_R}$ *for*

$$R := \frac{\mathrm{Lip}(r)}{c}+\frac{\mathrm{Lip}\bigl((\sigma\sigma^{\mathsf{T}})^{-1}\bigr)K^2}{2\sqrt{c}}\,,$$

where K *is the smallest positive root (using* (B3) (ii)*) of*

$$\frac{\sqrt{c}}{2}\,\|\sigma\sigma^{\mathsf{T}}\|_\infty\,\mathrm{Lip}\bigl((\sigma\sigma^{\mathsf{T}})^{-1}\bigr)\,x^2-c^{5/4}x+\mathrm{Lip}(r)\|\sigma\sigma^{\mathsf{T}}\|_\infty=0\,.$$

For the stochastic differential game in (4.2) we consider the following relative value iteration equation:

$$\begin{aligned}\frac{\partial\varphi}{\partial t}(t,x) &= \min_{v\in V}\,\max_{w\in\tilde{V}}\,\Bigl[\tilde{L}\varphi(t,x,w,v)+r(x,v)-\tfrac{1}{2}\,w^{\mathsf{T}}\bigl(a^{-1}(x)\bigr)w\Bigr]\;-\;\varphi(t,0)\,,\\ \varphi(0,x) &= \varphi_0(x)\,,\end{aligned}$$

where $\varphi_0\in C_{\mathcal{V}}(\mathbb{R}^d)\cap C^2(\mathbb{R}^d)$ with

$$\mathcal{V}(x) \;=\; \frac{(x^{\mathsf{T}}Qx)^{1+\alpha}}{\varepsilon+(x^{\mathsf{T}}Qx)^{1/2}}\,,$$

for some positive constants ε and α. Here note that Assumption (B3) implies Assumption (A3) of Section 2 for the Lyapunov function $\mathcal{V}$ given above, see [2, equation (7.3.6), p. 257].

By Theorems 3.1 and 4.1 the following holds.

Theorem 4.2. *Assume* (B1)–(B3). *For each* $\varphi_0\in C_{\mathcal{V}}(\mathbb{R}^d)\cap C^2(\mathbb{R}^d)$, $\varphi(t,x)$ *converges to* $\varphi^*(x)+\beta$ *as* $t\to\infty$.

The relative value iteration equation for the risk-sensitive control problem is given by

$$\begin{aligned}\frac{\partial\psi}{\partial t}(t,x) &= \min_{v\in V}\,\bigl[L\psi(t,x,v)+(r(x,v)-\ln\psi(t,0))\psi(t,x)\bigr]\,,\\ \psi(0,x) &= \psi_0(x)\,,\end{aligned}\qquad (4.3)$$

where

$$Lf(x,v) := b(x,v)\cdot\nabla f(x) + \frac{1}{2}\,\mathrm{tr}\bigl(a(x)\nabla^2 f(x)\bigr)\,, \quad f\in C^2(\mathbb{R}^d)\,.$$

That one has $\ln\psi(t,0)$ instead of $\psi(t,0)$ as the 'offset' is only natural, because we are trying to approximate the *logarithmic* growth rate of the cost. We have the following theorem:

Theorem 4.3. *Let ψ^* be the unique solution in the class of functions which grow no faster than $\mathrm{e}^{\|x\|^2}$ of the HJB equation for the risk-sensitive control problem given by*

$$\beta\psi^* \;=\; \min_{v\in V}\,\bigl[L\psi^*(x,v) + r(x,v)\psi^*\bigr]\,, \quad \psi^*(0)=1\,.$$

Under assumptions (B1)–(B3) *the solution $\psi(t,x)$ of the relative value iteration in* (4.3) *converges as $t\to\infty$ to $\mathrm{e}^{\beta}\psi^*(x)$ where β is the value of the risk-sensitive control problem given in Theorem 4.1.*

Proof. A straightforward calculation shows that $\psi^* = \mathrm{e}^{\varphi^*}$, where φ^* is given in Theorem 4.1. Then it easily follows that $\psi(t,x) = \mathrm{e}^{\varphi(t,x)}$, where φ is the solution of the relative value iteration for the stochastic differential game in (4.2). From Theorem 4.2, it follows that $\psi(t,x)\to\mathrm{e}^{\beta}\psi^*(x)$ as $t\to\infty$, which establishes the claim. □

5. Acknowledgement

The work of Ari Arapostathis was supported in part by the Office of Naval Research under the Electric Ship Research and Development Consortium. The work of Vivek Borkar was supported in part by Grant #11IRCCSG014 from IRCC, IIT, Mumbai.

References

[1] A. Arapostathis and V. S. Borkar, A relative value iteration algorithm for non-degenerate controlled diffusions, *SIAM J. Control Optim.*, vol. 50, No. 4, pp. 1886–1902, 2012.

[2] A. Arapostathis, V. S. Borkar and M. K. Ghosh, *Ergodic Control of Diffusion Processes*, Encyclopedia of Mathematics and its Applications **143**, Cambridge University Press, Cambridge, UK, 2012.

[3] V. E. Beneš, Existence of optimal strategies based on a specified information, for a class of stochastic decision problems, *SIAM J. Control*, vol. 8, pp. 179–188, 1970.

[4] V. S. Borkar and M. K. Ghosh, Stochastic differential games: Occupation measure based approach, *J. Optim. Theory Appl.*, vol. 73, No. 2, pp. 359–385, 1992, Errata: vol. 88, No. 1, pp. 251–252, 1996.

[5] V. S. Borkar and K. Suresh Kumar, Singular perturbations in risk-sensitive stochastic control, *SIAM J. Control Optim.*, vol. 48, No. 6, pp. 3675–3697, 2010.

[6] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics in Mathematics, Springer, Reprint of the 1998 Edition.

[7] D. J. White, Dynamic programming, Markov chains, and the method of successive approximations, *J. Math. Anal. Appl.*, vol. 6, No. 3, pp. 373–376, 1963.

Department of Electrical and Computer Engineering, The University of Texas at Austin, 1 University Station, Austin, TX 78712
E-mail address: ari@mail.utexas.edu

Department of Electrical Engineering, Indian Institute of Technology, Powai, Mumbai 400076, India
E-mail address: borkar.vs@gmail.com

Department of Mathematics, Indian Institute of Technology, Powai, Mumbai 400076, India
E-mail address: suresh@math.iitb.ac.in